Research Statement

Xuewen Yu

My research areas are Econometric Theory and Applied Econometrics. My research is motivated by the features of real world data like nonstationarity, instability, heterogeneity, high dimensionality, nonlinearity, persistency, and thus my goal is to develop econometric estimation, inference, and forecasting tools that accommodate such features in the data generating process.

To be more specific, my past and ongoing projects span several areas: (1). Nonstationary time series: issues around structural breaks in both mean and variance; (2). Large time-varying vector autoregressions; (3). Dynamic panel data models. In what follows, I elaborate upon each of them in turn. The mentioned papers (and associated code) are available at www.xuewenyu.com.

Area #1: Nonstationary time series: issues around structural breaks in both mean and variance.

[Paper 1] Bootstrap procedures for detecting multiple persistence shifts in heteroskedastic time series, with Mohitosh Kejriwal and Pierre Perron, 2020, *Journal of Time Series Analysis*, 41, 676-690.

[Paper 2] A two-step procedure for testing partial parameter stability in cointegrated regression models, with Mohitosh Kejriwal and Pierre Perron, 2021, forthcoming at *Journal of Time Series Analysis*.

[Paper 3] Generalized forecast averaging in autoregressions with a near unit root, with Mohitosh Kejriwal, 2021, *The Econometrics Journal*, 24, 83-102.

[Paper 4] Inference in mildly explosive autoregressions under unconditional heteroskedasticity, with Mohitosh Kejriwal, 2021, working paper.

Autoregressions are fundamental tools for analyzing economic time series. In general, we classify them into two paradigms—stationary (autoregressive parameter smaller than 1, integrated of order 0, I(0) hereafter) and nonstationary (autoregressive parameter equal to 1, I(1) hereafter, or greater than 1, I(e) hereafter) due to different theoretical foundations. While it is more natural to believe that common economic time series are a hybrid of both worlds, it is difficult to distinguish them rigorously, especially in view of potentially time-varying volatility. In determining multiple interchanges between I(0) and I(1), we propose a robust bootstrap procedure to inference under very mild assumptions (**Paper 1**). The new bootstrap procedures developed in this paper allow users to identify switches in persistence while controlling for various types of mean and variance instabilities, which, if not accounted for, will result in misleading inference. We apply our method to inflation rate data of OECD countries and find less evidence of breaks in persistence after controlling for the effects of changing mean and volatility.

In **Paper 2**, we consider the problem of testing for structural breaks in bivariate nonstationary time series regressions which are so-called cointegrated regressions. We show that the empirical and asymptotic sizes of existing tests are sensitive to the magnitude of breaks in parameters not under test. To remedy this problem, we propose and theoretically justify a simple two-step procedure for testing the stability of any subset of model parameters. The modeling setup is quite general and can accommodate both I(1) and I(0) regressors. In an application to US money demand, we show how the proposed approach can be fruitfully employed to estimate the welfare cost of inflation.

My third paper in this area focuses on inference in mildly-explosive autoregression, which is robust to time-varying volatility (heteroskedasticity) and autocorrelation in the innovations (**Paper 4**). We answer a simple yet unexplored question in the nonstationary time series literature: is heteroscedasticity and autocorrelation consistent (HAC) inference robust enough to cover the explosive autoregression case? Conventional wisdom tells us that HAC inference works well in the stationary world but usually fails in the nonstationary world. In contrast, our paper shows HAC inference continues to be valid in the mildly explosive autoregression. The resulting simple standard normal inference can guide users to analyze explosive behavior, such as financial bubbles, with the advantage of being agnostic about the underlying heteroskedasticity and serial correlation conveniently. As a side result, we also present evidence that the popular fixed-b approach is invalid under (unconditional) heteroskedasticity. We revisit an example of analyzing several representative stock indices by providing evidence of heteroskedasticity and constructing more informative confidence intervals of the explosive intensity compared to existing methods.

The three aforementioned papers deal with the problem of inference in nonstationary time series models. When the objective is forecasting, a practical approach is to combine stationary and nonstationary forecasts instead of pretesting. **Paper 3** proposes a generalized model averaging predictor for persistent time series in the presence of persistence uncertainty (I(0) or I(1)) and lag order uncertainty. Specifically, we provide theoretical contributions for generalized averaging over the stationary and the nonstationary predictors in the presence of potential trend estimation uncertainty and lag order uncertainty. In an application to US macroeconomic time series, we illustrate the efficacy of the advocated method in practice and find that both persistence and lag order uncertainty have important implications for the accuracy of forecasts.

Going forward: My ongoing research in this area extends and modifies the idea in my **Paper 1** to financial bubble detection and surveillance, which are crucial issues to practitioners and policymakers nowadays. Essentially, bubble testing can be cast as a structural break detection problem. More specifically, the problem is to detect whether the autoregressive parameter switches between I(1) and I(e) regimes in the potential presence of time-varying volatility. In a joint project with Mohitosh Kejriwal and Pierre Perron, "Testing for Multiple Bubbles under Nonstationary Volatility", we employ adaptive estimation in conjunction with a wild bootstrap procedure to detect bubbles under time-varying volatility. Through extensive simulations, we find our method provides adequate size and at the same time offers monotonic and substantially higher power than extant workhorse methods which allow for heteroskedasticity. These features should be of relevance to empiricists when testing for and monitoring bubbles in economic and financial time series.

A second ongoing project is "Multistep Forecast Averaging with Stochastic and Deterministic Trends", joint with Mohitosh Kejriwal and Linh Nguyen, which extends the one-step ahead setup of my **Paper 3** to a multi-step scenario and presents a new approach to forecast combination in the presence of uncertainty about the nature of persistence, lag order and the underlying trends.

Area #2: Large time-varying vector autoregressions (VARs).

[Paper 5] Fast and accurate variational inference for large Bayesian VARs with stochastic volatility, with Joshua Chan, 2020, *under revision*.

[Paper 6] Large order-invariant Bayesian VARs with stochastic volatility, with Joshua Chan and Gary Koop, 2021, *under review*.

[Paper 7] Bayesian VARs with factor stochastic volatility: Identification, order invariance and model comparison, with Joshua Chan and Eric Eisenstat, 2021, working paper.

My Area # 1 mainly deals with issues around robust inference in low dimensional time series regressions. In Area # 2, my projects focus on multivariate and potentially high-dimensional autoregressions. High-dimensional framework is becoming increasingly popular in economic analysis. The big data era calls for econometricians to create large-scale models to utilize the massive data with the aim to sharpen econometric inference. Large VARs in particular have become increasingly popular among macroeconomic researchers for forecasting and structural analysis, and it is a fast-growing area that attracts both the theorists and empiricists' attention. One main obstacle of routinely using large VARs with flexible features such as time-varying volatility is that they are computationally intensive to estimate. In a joint work with Joshua Chan, we develop a fast and accurate procedure to estimate high-dimensional VARs using variational inference which originally comes from the machine learning literature (**Paper 5**). The striking advantage of using variational inference is that it makes handling high dimensions, such as 100 variables in a system at an exceptionally fast rate which is usually impractical for traditional Markov Chain Monte Carlo (MCMC) methods. Moreover, our variational Bayes algorithm is globally optimized, which has been shown to behave more accurately than other approximation methods. The tools provided in our paper are beneficial for studying time-varying high dimensional time series problems in many areas. In our paper, we provide an example of a 96-variable VAR with stochastic volatility to measure global bank network connectedness, which is an important issue in modern financial risk management.

My second and third papers in this area focus on large order-invariant VARs. Structural VARs with Cholesky stochastic volatility has been the building block models for macroeconometric forecasting and inference over the past twenty years. However, a well-known drawback of the Cholesky stochastic volatility model is that it is not not invariant to the way the variables are ordered. Moreover, the ordering issues become more severe as the dimension grows. To tackle this problem, one of my papers utilizes the role of stochastic volatility to achieve identification and then proposes an efficient MCMC sampler for an order invariant model (**Paper 6**). Specifically, the usual setup assumes a lower triangular structure to achieve identification. We showed that that is not necessary. In our model, it is a full and identified (up to permutation and sign changes) structure that is guaranteed to be invariant to ordering. In short, we contribute to the literature by relaxing the lower triangular constraints of the baseline model and achieving a totally full structure. Through various simulations and empirical investigations, we confirm that the performance of the baseline model differs considerably among different order specifications, especially in terms of likelihood. On the contrary, our order invariant model performs quite stable across different orders, and the full structure is strongly supported by the real data. In a forecasting exercise using monthly macroeconomic data, we find that the forecasting performance compared to the baseline model can be significantly improved by allowing a full matrix structure.

My third paper develops another order invariant model and focuses on structural analysis (instead of forecasting). The model is designed to allow for a large number of sign restrictions to be imposed in the estimation. Specifically, we construct a VAR in which the innovations have a factor structure, and both the factors and the idiosyncratic errors follow independent stochastic volatility processes (**Paper 7**). We contribute theoretically that the likelihood implied by this model is invariant to the orders, and the factor loadings are identifiable provided with several sufficient conditions we make. We show how the usual structural analysis tools for the new model like impulse response analysis and various decompositions can be implemented, which are similar to those designed for standard structural VARs. In summary, our method provides a way to conduct structural analysis that scales well in the big data era.

Going forward: I have a few ongoing large Bayesian VAR projects. For example, one project extends the variational Bayes approach in my **Paper 5** to tackle the problem of comparing different shrinkage priors. Starting from the recent literature, more flexible shrinkage priors, like various global-local priors, are adopted for large VARs, though which one is the best fit remains ambiguous since computing their marginal likelihood is generally intractable. In a joint work with Joshua Chan and Wei Zhang, "Comparing Shrinkage Prior Specifications for Large Bayesian VARs by Variational Bayes", we provide a fast approximation method for model comparison by variational Bayes.

Following my **Paper 6**, due to space limitations, only forecasting is discussed with the new model. Our next step is to develop structural inference tools under the new model and analyze the properties compared to the baseline model. There is evidence drawn in the literature that the structural inference results are sensitive to variable ordering, and even they could be opposite with certain real macroeconomic data when the orders are reversed. It would be interesting to study such scenarios with our new model to retrieve a unique conclusion. On the other hand, the ingredient that lower triangular structure is relaxed in my **Paper 6** opens a door to many models that assume the same Cholesky structure, such as time-varying parameter VARs with constant volatility, regime switching VARs, threshold VARs, etc. It would be very useful to extend those models to order invariant ones that are more general.

Area #3: Dynamic panel data models.

[Paper 8] Indirect inference estimation of dynamic panel data models, with Yong Bao, 2021, revise and resubmit requested by *Journal of Econometrics*.

So far, my research portfolio deals with regressions in the pure time series aspect. In a panel autoregression (i.e., dynamic panel data model) setting, joint with Yong Bao, we propose an estimator for higher-order dynamic panel models based on the idea of indirect inference by matching the simple within-group estimator with its analytical approximate expectation (**Paper 8**). The resulting estimator is shown to be consistent and asymptotically normal. Monte Carlo simulations show that the proposed estimator is virtually unbiased, achieves usually lower root mean squared error than the generalized method of moments (GMM) estimator, and delivers very reliable empirical size across various parameter configurations and error distributions. As a side result, for the special case of first-order dynamic panel, we correct a widely-used estimator in the dynamic panel literature whose idea was similarly based on bias correction. In a practical sense, our estimator enjoys a relative simple form, and there is no need for constructing various instruments for estimation compared with the popular tools such as IV or GMM. These advantages together will be attractive to researchers to use our estimator in empirical analysis. The proposed estimator can be used to study problems in both empirical microeconomics or macroeconomics. In our paper, we provide an example of using the new estimator to estimate convergence parameter in inequality measures among 63 countries during 1985-2015. It shows strong evidence of convergence over long test horizons but much weaker evidence over a 5-year horizon for developing countries.

Going forward: The neat formula of the estimator provided in our paper fascinates me to further think about finding similar estimators in other areas of dynamic panel data literature. For example, in our paper, we only focus on linear model settings. In a nonlinear dynamic panel, it would be interesting to extend our indirect inference method to bias correct the bias or the likelihood. Such an example could be the dynamic discrete choice model, a classical and popular tool used by microeconomic researchers. Another example will be the quantile dynamic panel model. It would be interesting to extend our indirect inference method there to present an new solution and investigate its advantages over the existing solutions provided in the literature.

A second interesting issue is that in a dynamic panel setting, the effect of uncertain order of lags is non-negligible for estimation, inference and forecasting. A natural extension of our paper will be studying averaging estimators in the panel autoregression setup using frequentist model averaging methodologies (in my **Paper 3**) along with our indirect inference estimator.

Lastly, an ongoing project extended from this paper is joint with Yong Bao, "A Test of Test Horizon in Convergence Studies", which is also under a panel autoregression setup, but with a non-consecutive lag component. This model is motivated by development economists who wish to test whether the income distribution (say, the Gini coefficient) among different countries converges at a specific horizon. In practice, it is common that the researchers face an uncertain test horizon. In this project, we try to provide guidance to the empiricist for finding the appropriate time horizon over a variety of choices. Interestingly, the method we find relates to some ideas in the structural break literature which are adopted in my Area #1. Through several simulations, the sup-Wald test borrowed from the structural break literature works well in pinning down the true horizon.